CSE 2194: Supervised Machine Learning
Programming Assignment-V
(Linear Regression, Logistic Regression)

Part -1 (Simple Linear Regression)
Question 1:
Download the dataset
1. Print the head of the data frame.
2. Analyze the data by looking at the data types.
3. Create plots to understand more about the distribution.
4. Separate the dependent and the independent variables and store them in newly defined variables.
5. Split the data into 25% test and 75% train.
6. re-shape the data since there is only one dependent and one independent variable, whereas the model.fit and model.predict functions always expect a 2-D array.
7. Train the linear regression model and predict the values for test set data.
8. Evaluate the built model using root mean squared error R2 score.

(Multiple linear regression)
Question 2:
Download the given dataset. This data input features Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, and Sample Question Papers Practiced. The corresponding output feature is the Performance Index. Write a Python program to perform the following tasks.

1. Exploring the data, cleaning it if required, and printing the head of the data.
2. Apply the preprocessing techniques for object columns.
3. Create plots to understand more about the distribution.
4. Separate the dependent and the independent variables and store them in newly defined variables.
5. Split the data into 30% test and 70% train.
6. Train the linear regression model and predict the values for test set data.
7. Evaluate the built model using root mean squared error R2 score.

Part 2 (Binary Logistic Regression)
Question 3:
Download the given dataset. Write a Python program to perform the following tasks.
1. Print the basic dataset information such as head, statistics and data info.
2. Create box plots to understand how different attributes are distributed for the Outcome variable.

3. Create pair plots of selected columns such as 'Glucose', 'Age', 'DiabetesPedigreeFunction', 'BMI', ' Insulin', 'SkinThickness', and 'Blood Pressure'.
4. Split the data into training and test data and fit our training data to a logistic regression model.
5. Create a confusion matrix.

(Multinomial Logistic Regression)
Question 4:

Download the given dataset.
1. Print the basic dataset information such as head, statistics and data info.
2. Apply the preprocessing techniques to convert the target columns into numerical columns.
3. Create pair plots for input features.
4. Split the data into training and test data and fit our training data to a logistic regression model.
5. Create a confusion matrix and accuracy score.